Accelerating Atmospheric Simulation on GPU, FPGA, and MIC

Haohuan Fu
haohuan@tsinghua.edu.cn
Center for Earth System Science
Tsinghua University, Beijing

Sep/19/2013 @ NCAR

The Center for Earth System Science, Tsinghua University

- Started in 2009

CMIP5: LASG-CESS

- Study of the earth as an integrated system
  - to investigate interactions between atmosphere, land, water, ice, biosphere, societies, technologies, and economics
  - observing, understanding, and predicting global changes
  - to guide political / economical / technical decisions at different scales for assuring sustainable development

The Present Faculty

Tsinghua HPGC Group

- HPGC: high performance geo-computing
  http://www.thuhpgc.org

- High performance computational solutions for geoscience applications
  - simulation-oriented research: providing highly efficient and highly scalable simulation applications (climate modeling, exploration geophysics)
  - data-oriented research: data processing, data compression, and data mining

- Combine optimizations from three different perspectives (Application, Algorithm, and Architecture), especially focused on new accelerator architectures

- Application
  - Climate Modeling
    - global-scale atmospheric simulation (800 Tflops Shallow Water Equation)
    - GPU-based acceleration for GEOS-Chem
    - FPGA-based acceleration for weather forecasting acceleration
  - Exploration Geophysics
    - forward modeling / inversion / migration
  - Remote Sensing Data Processing
    - data analysis, visualization, correlation of different data sets
- Algorithm
  - parallel Stencil on Different HPC Architectures
  - parallel Sparse Matrix Solver
  - parallel Data Compression (PLZMA)
  - hardware-Based Gaussian Mixture Model Clustering Engine: 517x speedup
- Architecture
  - multi-core/many-core (CPU, GPU, MIC)
  - reconfigurable hardware (FPGA)

- Application
  - Climate Modeling
    - global-scale atmospheric simulation (800 Tflops Shallow Water Equation)
    - GPU-based acceleration for GEOS-Chem
    - FPGA-based acceleration for weather forecasting acceleration
  - Exploration Geophysics
    - forward modeling / inversion / migration
  - Remote Sensing Data Processing
    - data analysis, visualization, correlation of different data sets
- Algorithm
  - parallel Stencil on Different HPC Architectures
  - parallel Sparse Matrix Solver
  - parallel Data Compression (PLZMA)
  - hardware-Based Gaussian Mixture Model Clustering Engine: 517x speedup
- Architecture
  - multi-core/many-core (CPU, GPU, MIC)
  - reconfigurable hardware (FPGA)

Outline

- Tianhe-1A: GPU
- Maxeler DFE: FPGA
- Tianhe-2: MIC
- Future Plan & Discussion

Multidisciplinary Collaborations

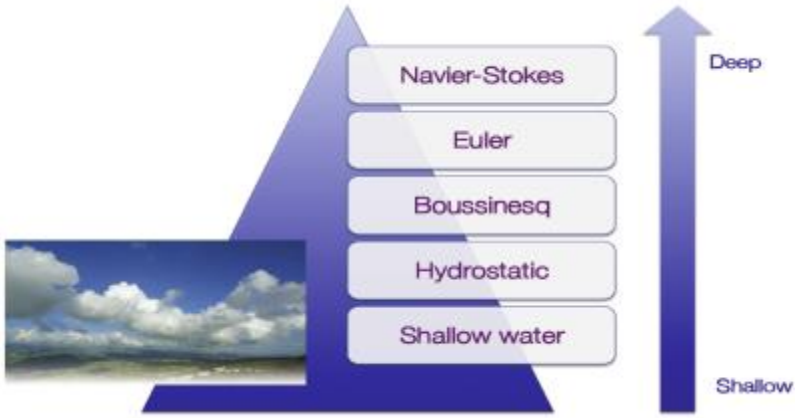
- Prof. Chao Yang
  - Institute of Software, CAS
  - computational mathematics
- Dr. Wei Xue
  - Department of Computer Science, Tsinghua University
  - HPC (MPI, OpenMP, MIC)
- Dr. Haohuan Fu
  - Center for Earth System Science, Tsinghua University
  - HPC (accelerators, GPU, FPGA, MIC)
- Prof. Lanning Wang
  - College of Global Change and Earth System Science, Beijing Normal University (BNU-ESM)
  - climate scientist

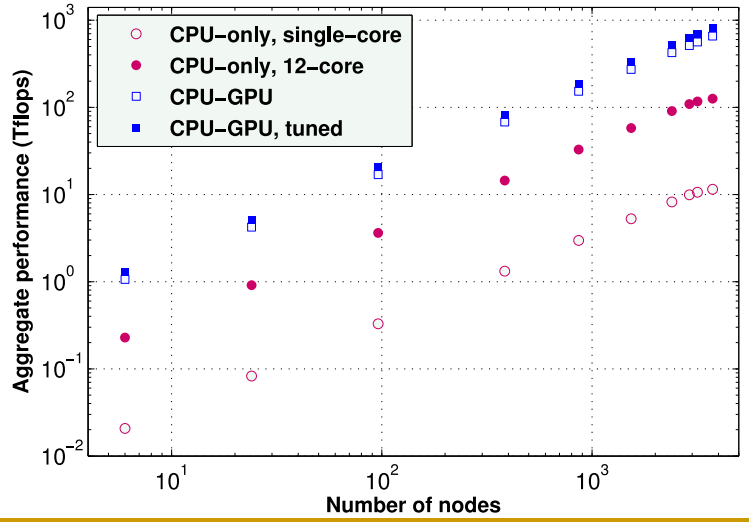
Outline

- Tianhe-1A: GPU
- Maxeler DFE: FPGA
- Tianhe-2: MIC
- Future Plan & Discussion

Highly-Scalable Framework for Global Atmospheric Simulation on Tianhe-1A

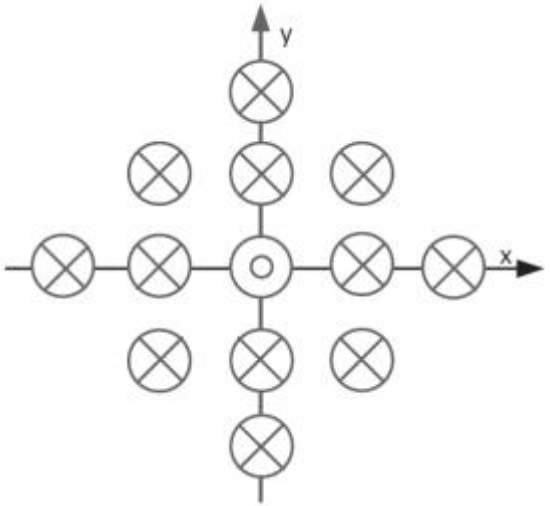
- Starting from shallow wave equation
  - cubed-sphere mesh grid
  - adjustable partition between CPU and GPU
  - scale to 40,000 CPU cores and 3750 GPUs with a sustainable performance of 800 TFlops



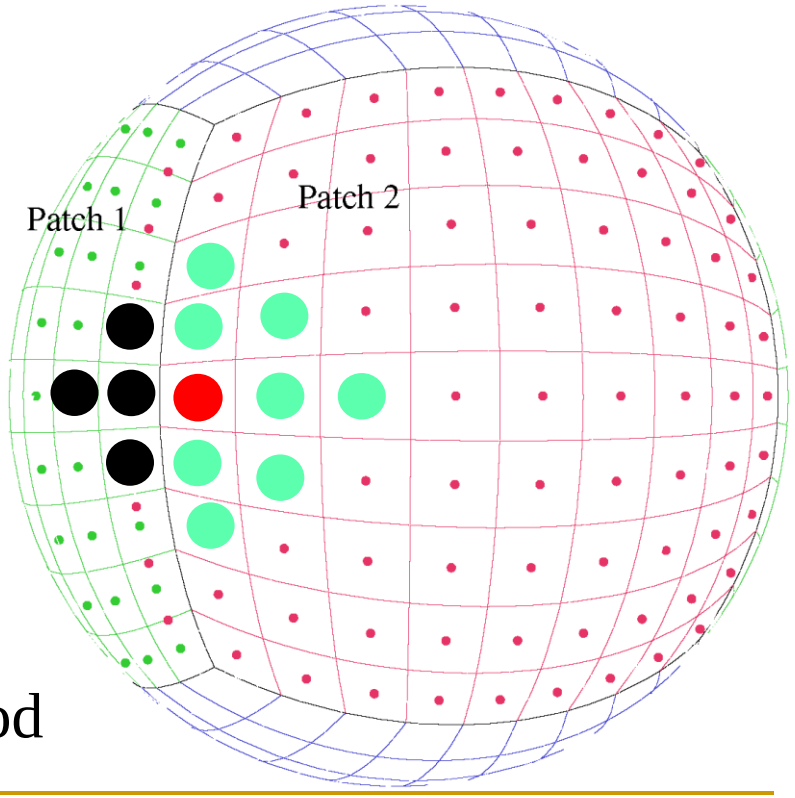


Mesh and Stencil

- 13-point stencil

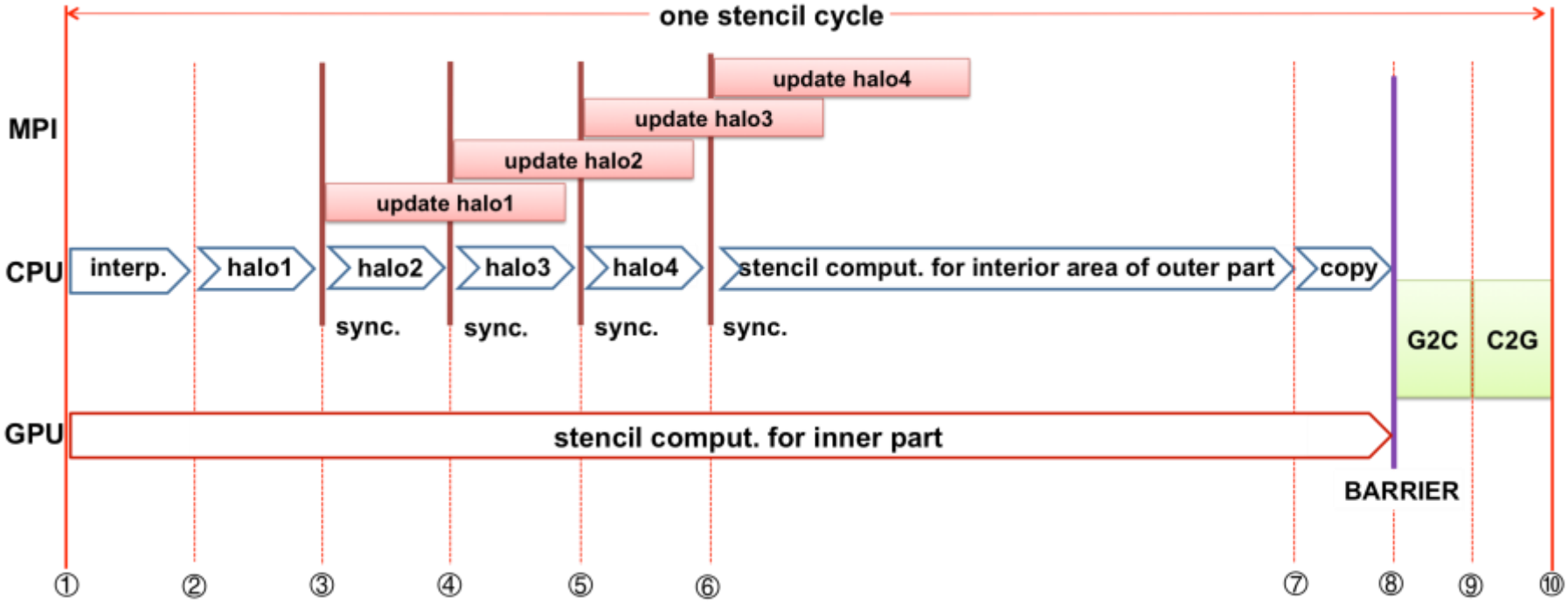


- Interp. Across patches
  - 1-d linear interpolation



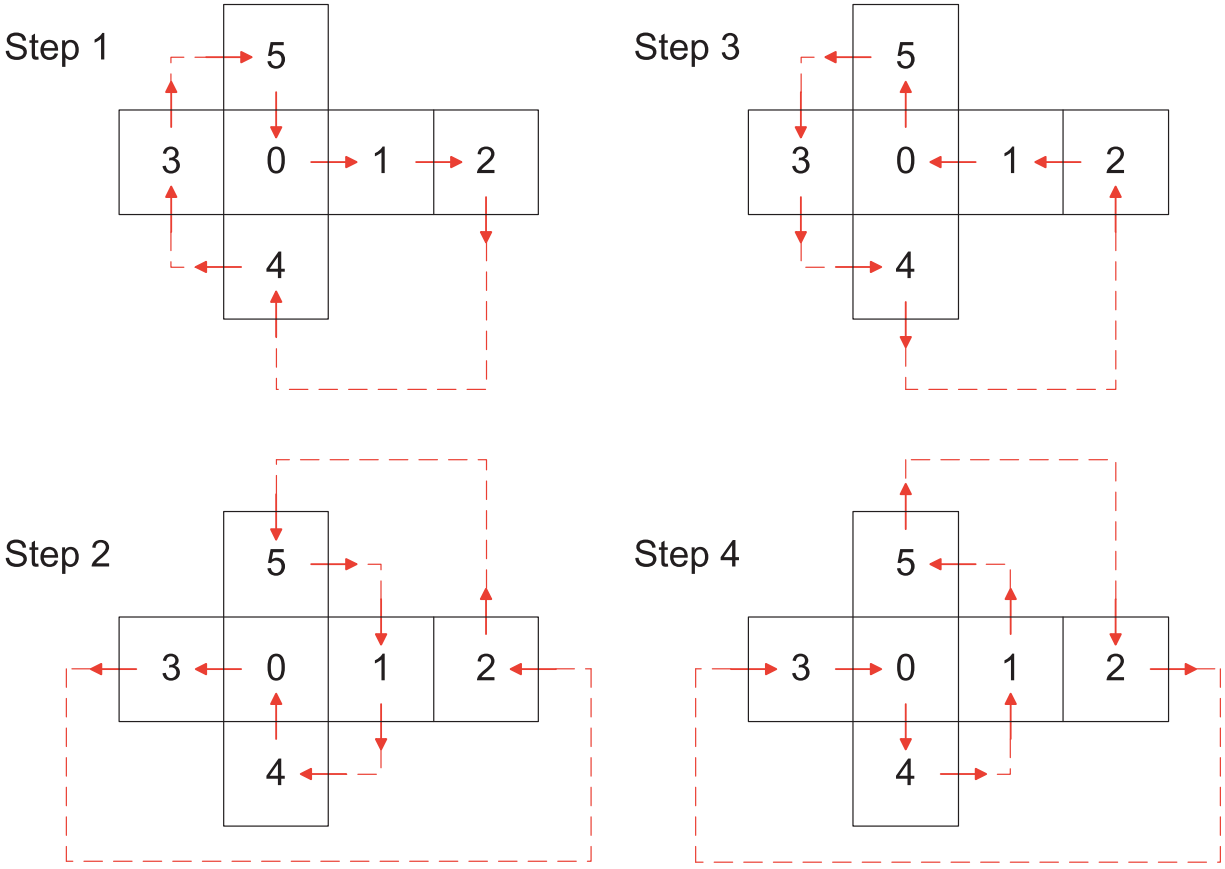
- Spatially discretized with a cell-centred finite volume method
- Integrated with a second-order accurate TVD Runge-Kutta method

Improved hybrid CPU-GPU Algorithm



Note: halo1/2/3/4 — the 4 steps of the “pipe-flow” communication scheme
adjustable partition between CPU and GPU

“Pipe-Flow” Scheme for Message-Passing on Cubed-Sphere



Four steps to arrange conflict-free message-passing on cubed-sphere.

The arrows indicate directions of data entering or exiting patches as a pipe flow.

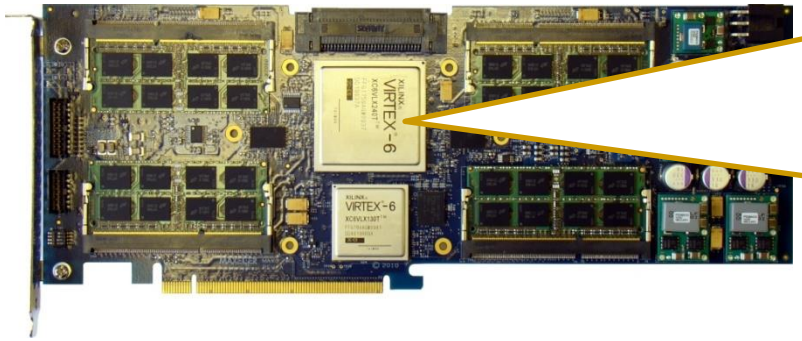
For more details, please refer to our PPoPP 2013 paper: “A Peta-Scalable CPU-GPU Algorithm for Global Atmospheric Simulations”, in *Proceedings of the 18th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (PPoPP)*, pp. 1-12, Shenzhen, 2013. .

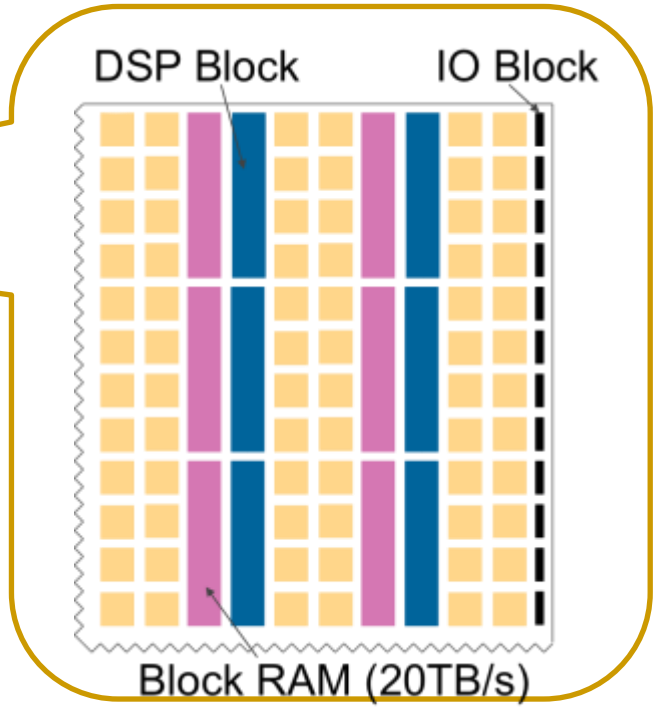
Outline

- Tianhe-1A: GPU
- Maxeler DFE: FPGA
- Tianhe-2: MIC
- Future Plan & Discussion

Highly-Scalable Atmospheric Simulation on Data-Flow Engines

- Maxeler Data-Flow Engine (DFE)
  - Field-Programmable Gate Arrays (FPGA)
  - 24 GB onboard memory
  - PCIE connection to host
  - MaxRing connection between cards





Hybrid CPU+FPGA Design

For each stencil cycle

FPGA side:

①Inner-part stencil

CPU side:

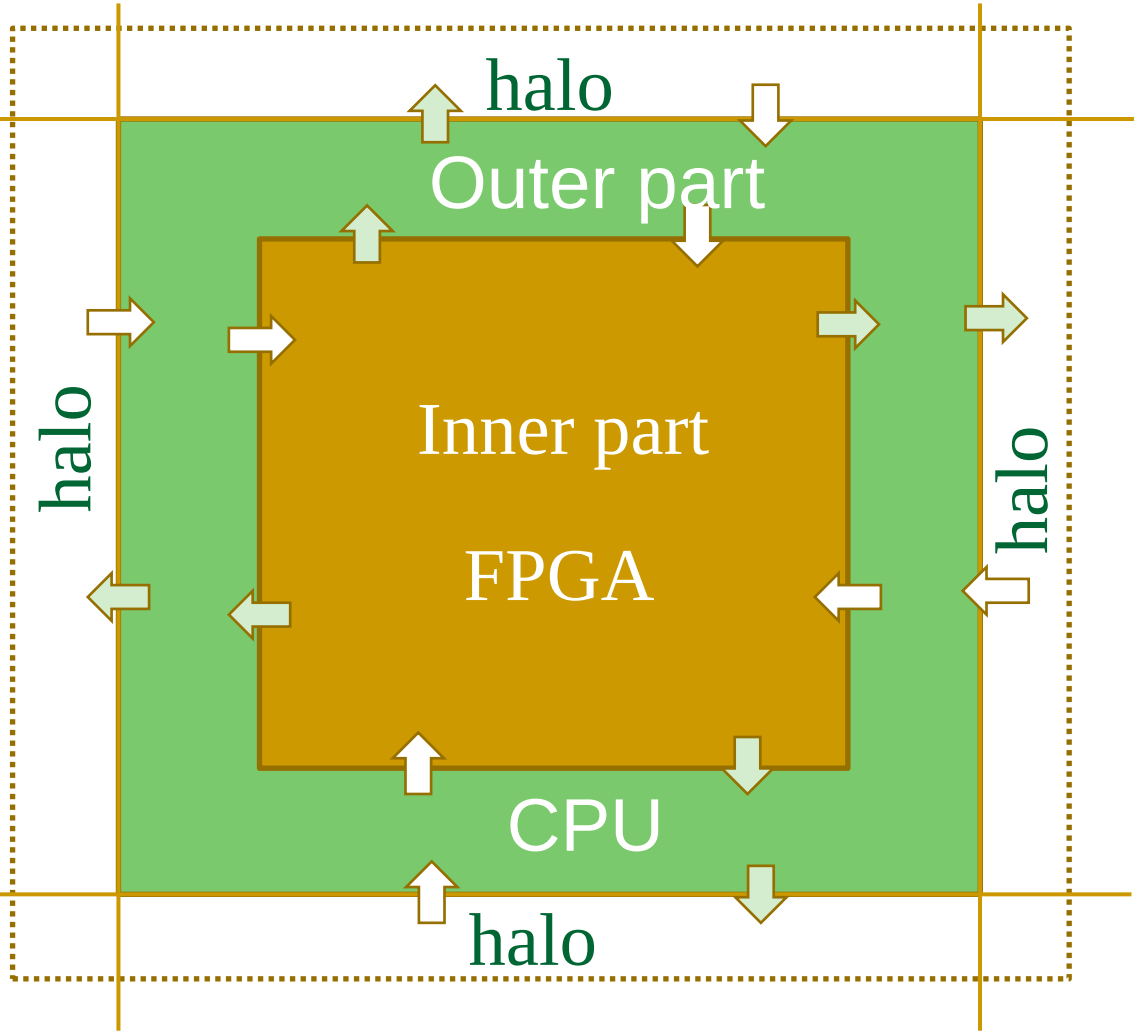
①Update halos

②Interpolate if necessary

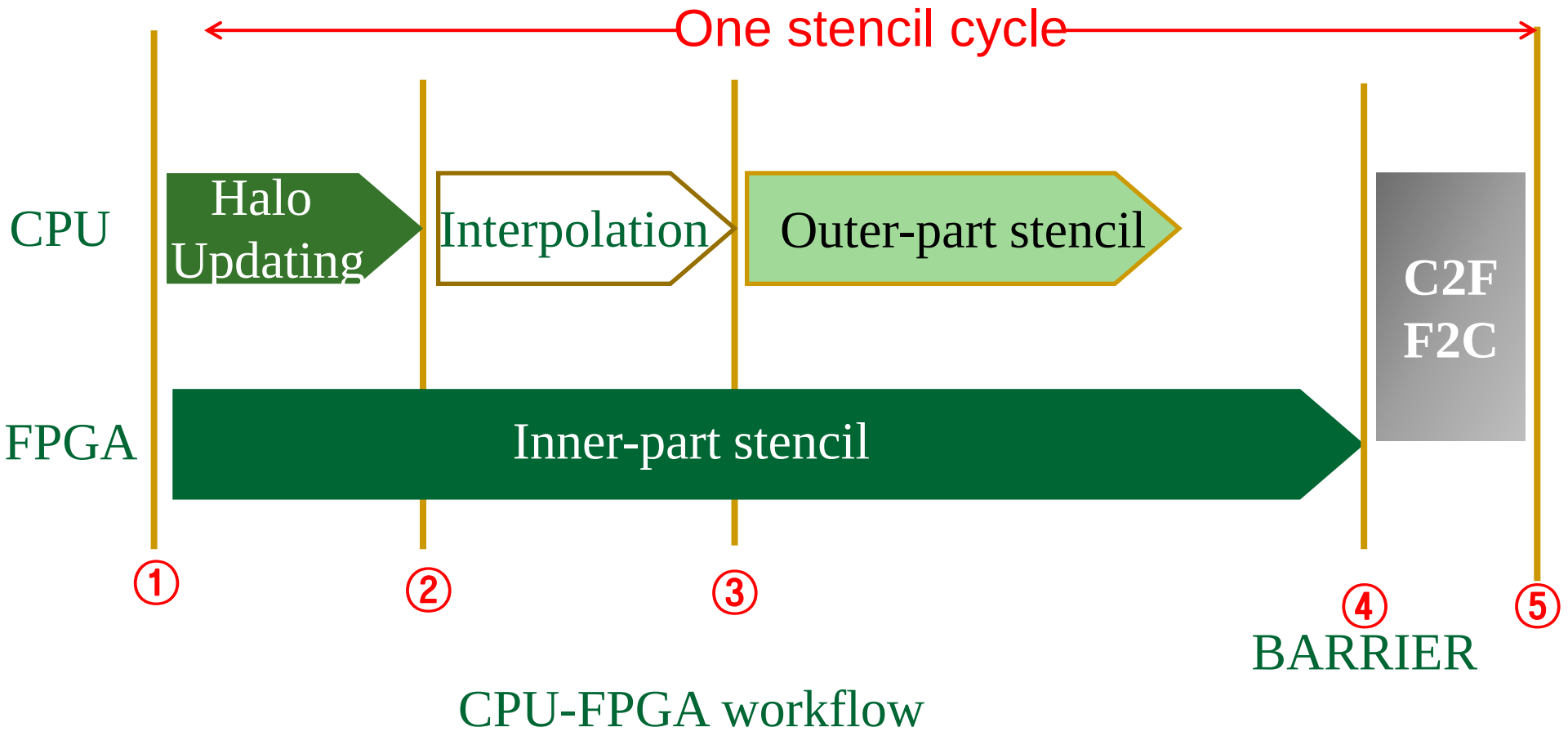
③Outer-part stencils

BARRIER:

CPU-FPGA exchange



Work Flow



CPU-FPGA workflow

Go for a Mixed-Precision Design

Floating point operations of SWE stencil

Operations	num
ADD/SUB	434
MUL	570
DIV	99
Others	45

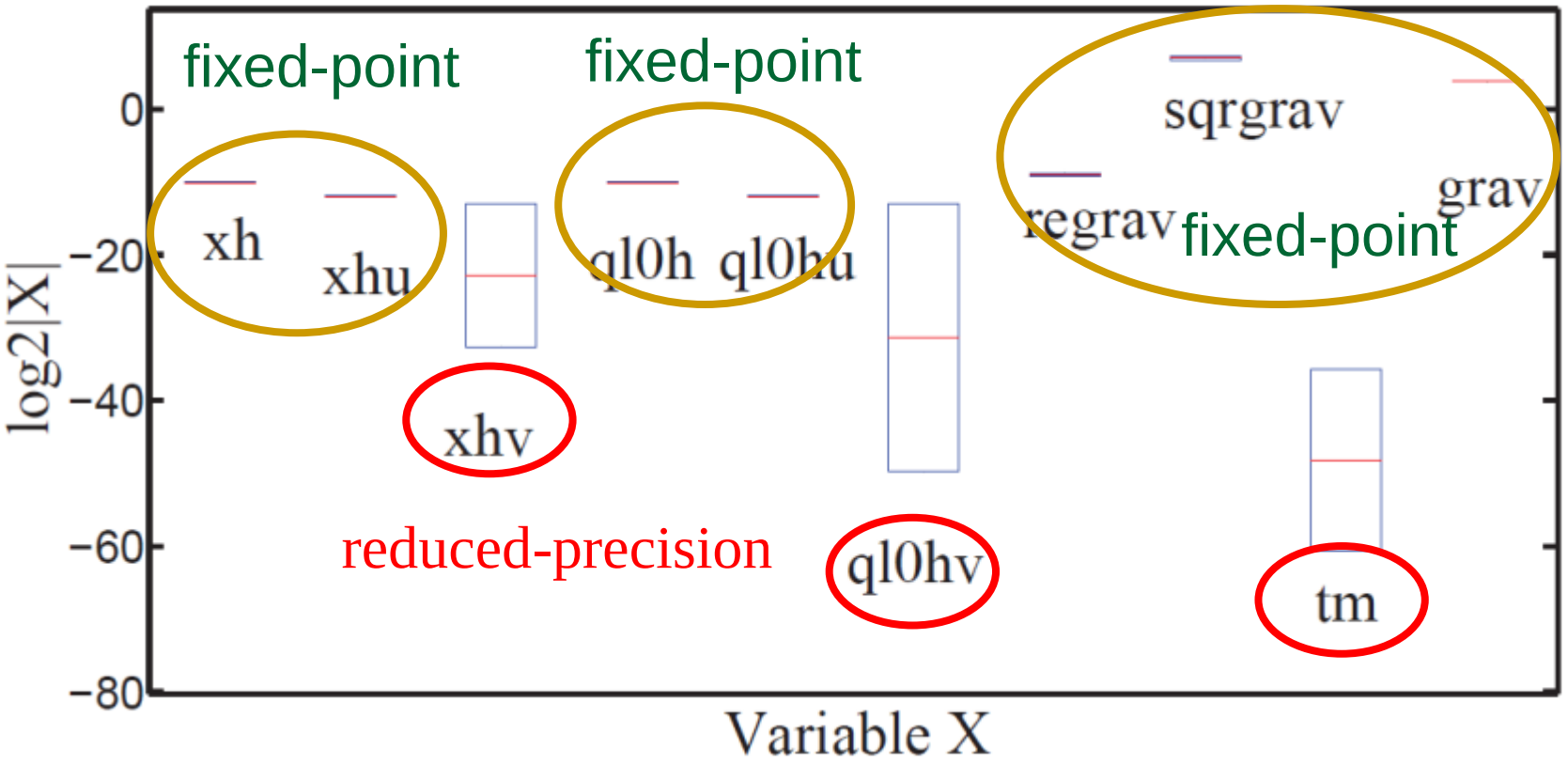
Resource Cost of SWEs on Virtex-6 SX457T

Resource	baseline
LUTs	299 %
FFs	220 %
BRAMs	20 %
DSPs	189 %

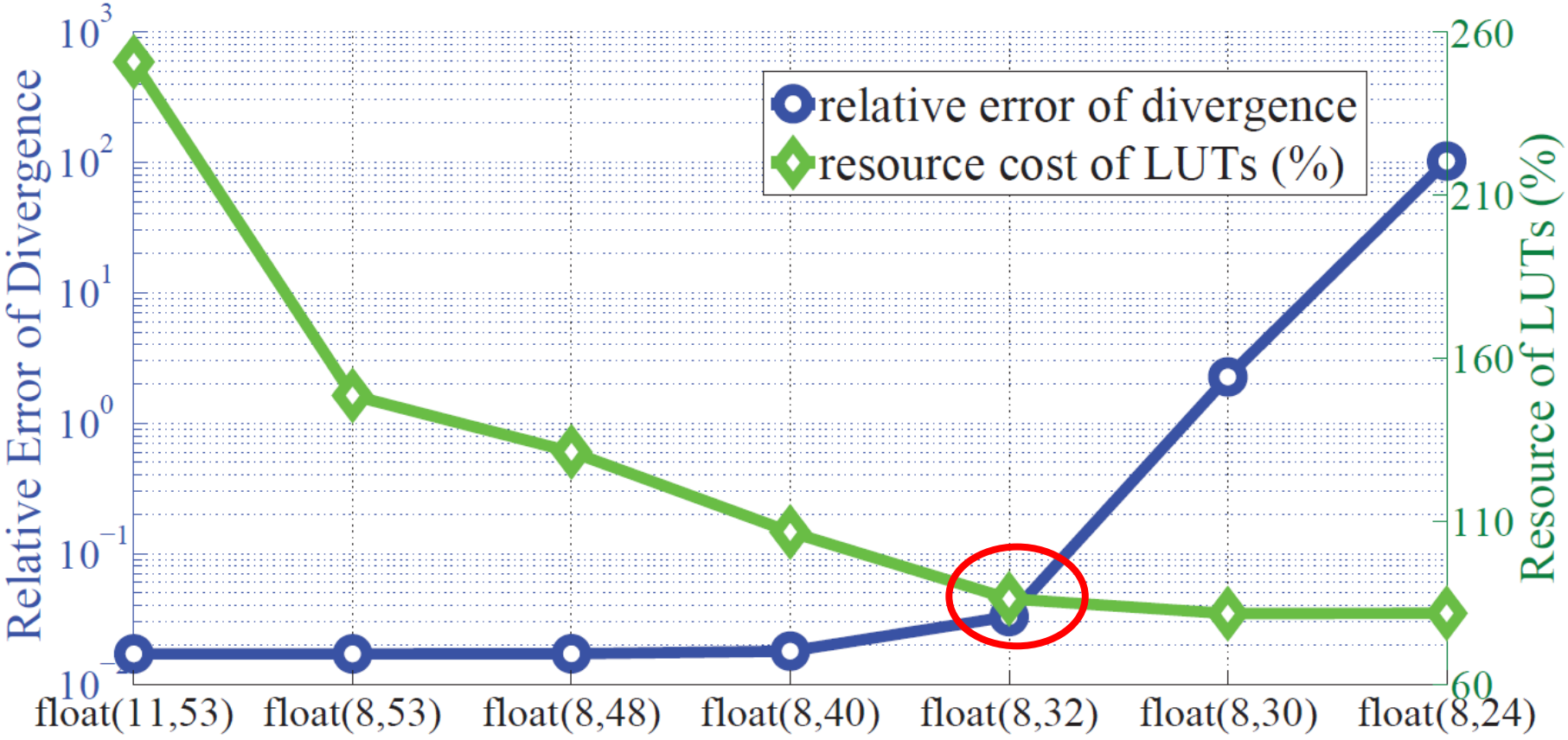
Baseline: a straightforward double-precision SWEs

- Precision-based optimization to further decrease the resource usage

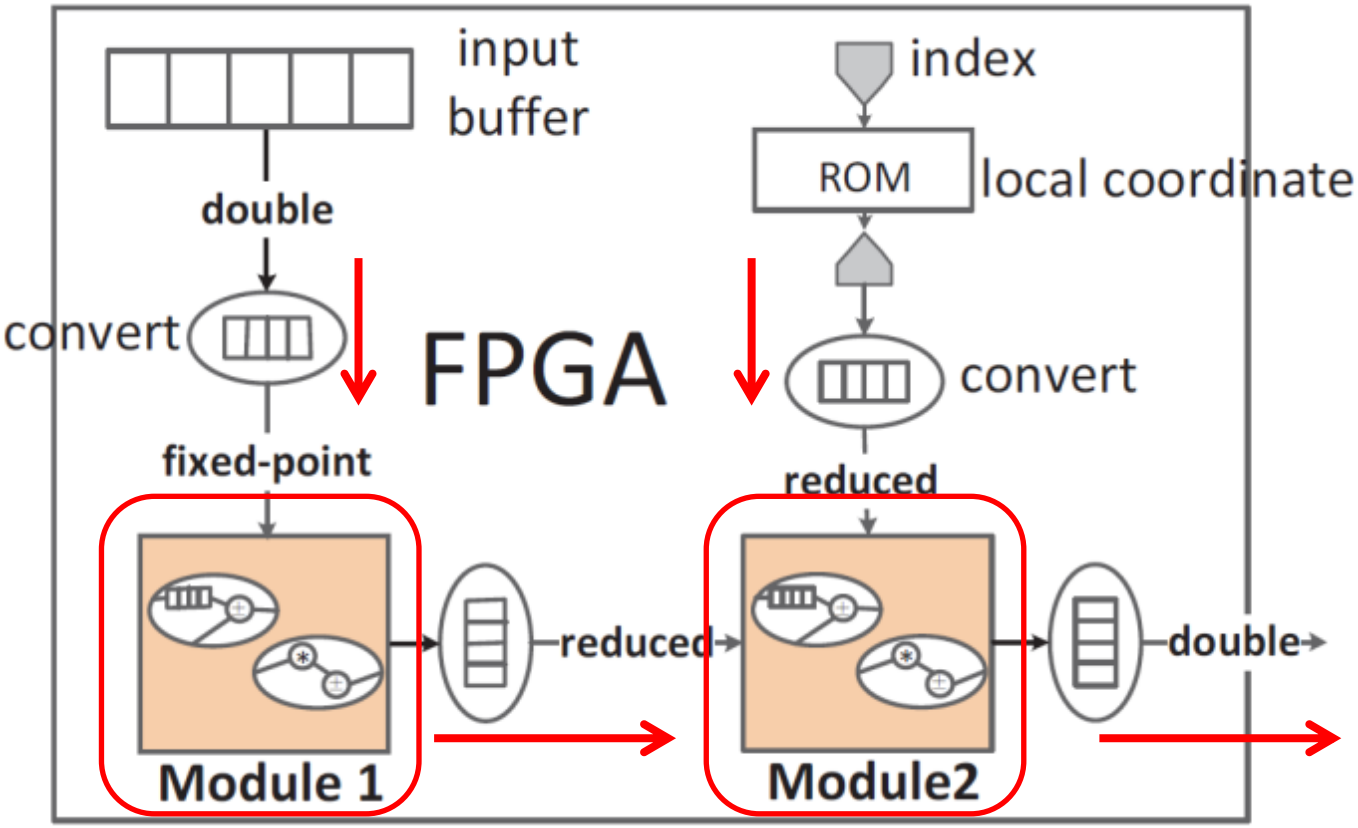
Analysis of the Dynamic Range



Precision Exploration



General Architecture of the Mixed-Precision Design



Resource Cost of SWEs on Virtex-6 SX457T

Resource	baseline	mixed-precision
LUTs	299 %	76.17%
FFs	220 %	53.41%
BRAMs	20 %	12.59 %
DSPs	189 %	44.84 %

- Baseline: a straightforward double-precision SWEs
- Mixed-precision: fixed-point and reduced-precision floating-point

Hardware Platform: Maxeler DFEs

- Environment
  - Maxcompiler development tool
- MaxWorkstation
  - One Intel i7 quad-core CPU
  - One Accelerator card (Virtex-6 SX 475T & 24 GB DRAM)
- MaxNode
  - 12 Intel Xeon CPU cores
  - four Accelerator cards (Virtex-6 SX 475T & 24 GB DRAM)

Performance Results

Platform	Performance $(points/second)$	Speedup
6-core CPU	4.66K	1
Tianhe-1A node	110.38K	23x
MaxWorkstation	468.1K	**100x**
MaxNode	1.54M	**330x**

14x

Meshsize: $1024 \times 1024 \times 6$

MaxNode speedup over Tianhe node: 14 times

Power Efficiency

Platform	Efficiency $(points/(second \times watt))$	Speedup
6-core CPU	20.71	1
Tianhe-1A node	306.6	14.8x
MaxWorkstation	2.52K	**121.6x**
MaxNode	3K	**144.9x**

9 x

Meshsize: $1024 \times 1024 \times 6$

MaxNode is 9 times more power efficient

For more details, please refer to our FPL 2013 paper: “Accelerating Solvers for Global Atmospheric Equations Through Mixed-Precision Data Flow Engine”, in *Proceedings of the 23rd International Conference on Field Programmable Logic and Applications*, 2013.

Outline

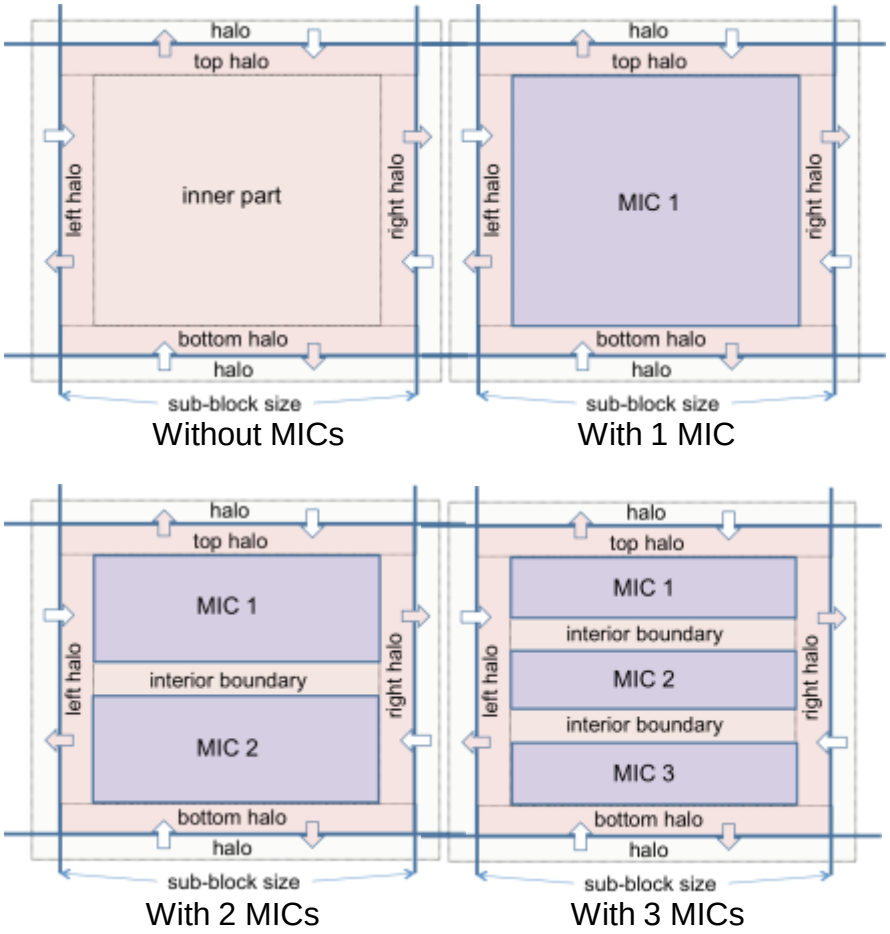
- Tianhe-1A: GPU
- Maxeler DFE: FPGA
- Tianhe-2: MIC
- Future Plan & Discussion

Tianhe-2: Brief Introduction

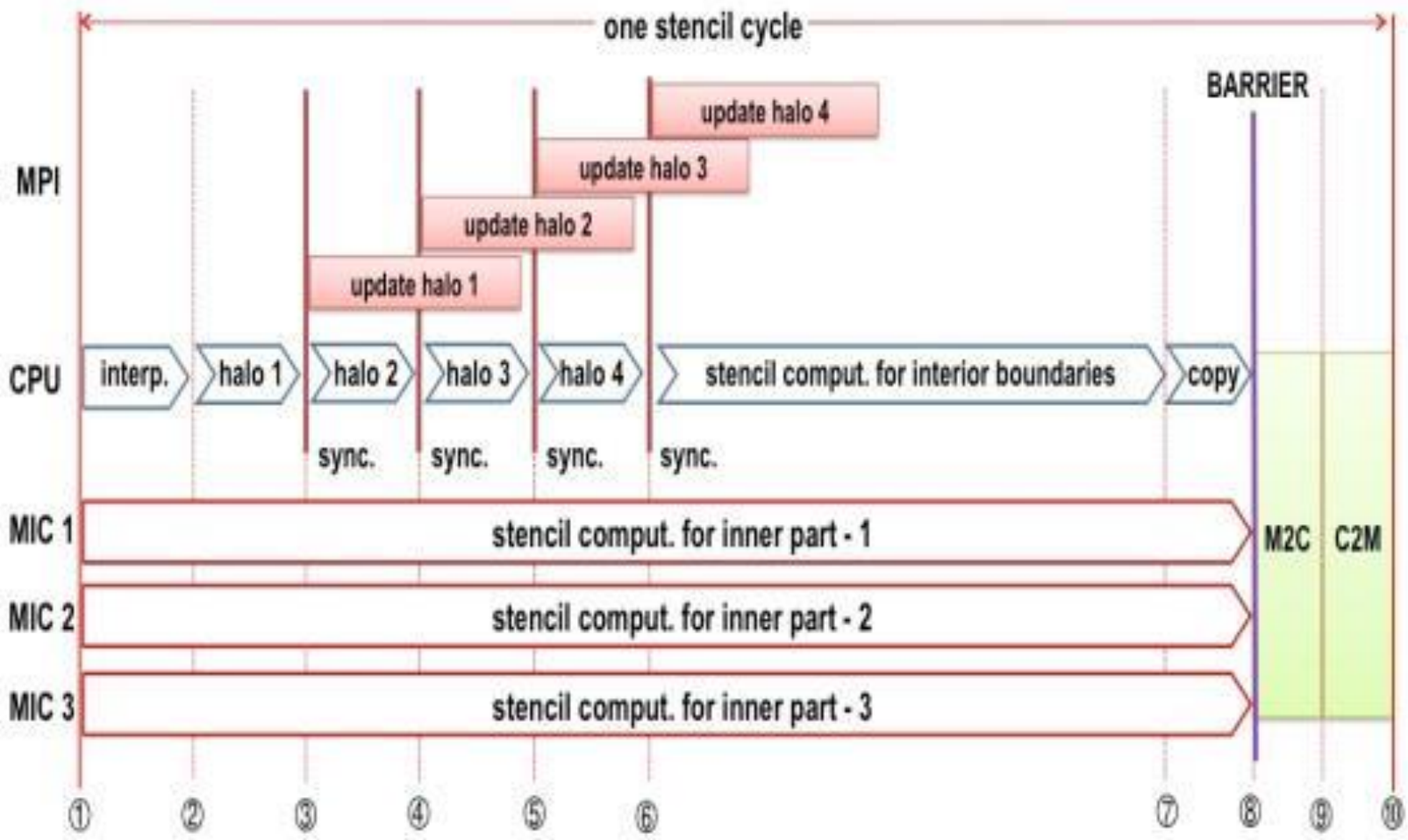
- Tianhe-2
  - 16,000 nodes
  - each node contains two 12-core Intel Ivy Bridge CPUs, and 3 Intel Xeon Phi Acceleration Cards
  - peak: 54.9 PFlops
  - LINPACK: 33.8 PFlops

Running SWE on Tianhe-2

- Hierarchical 2D domain decomposition
- Balanced CPU/MIC utilization
  - 0-3 MICs
  - adjustable blocks

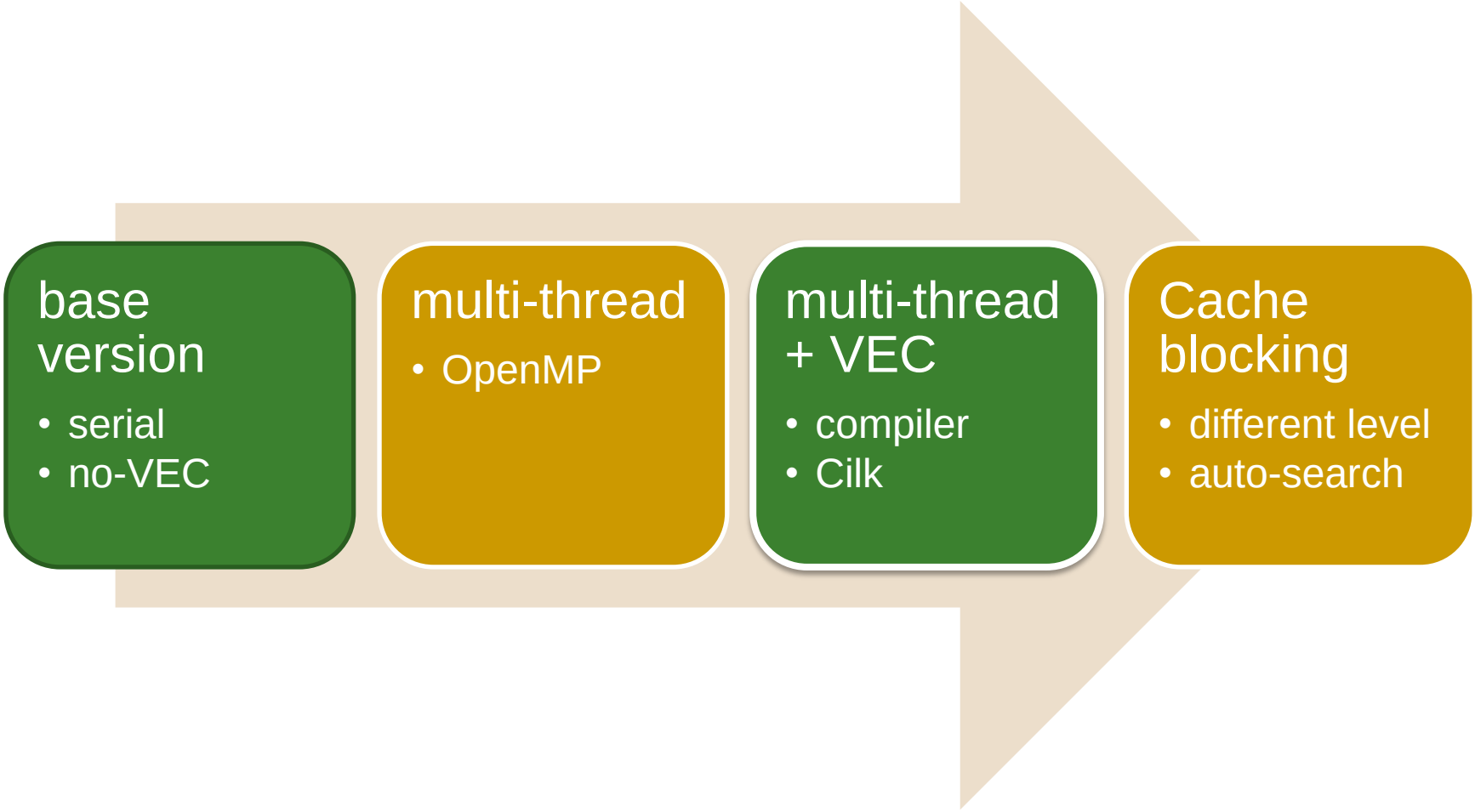


Running SWE on Tianhe-2: Workflow

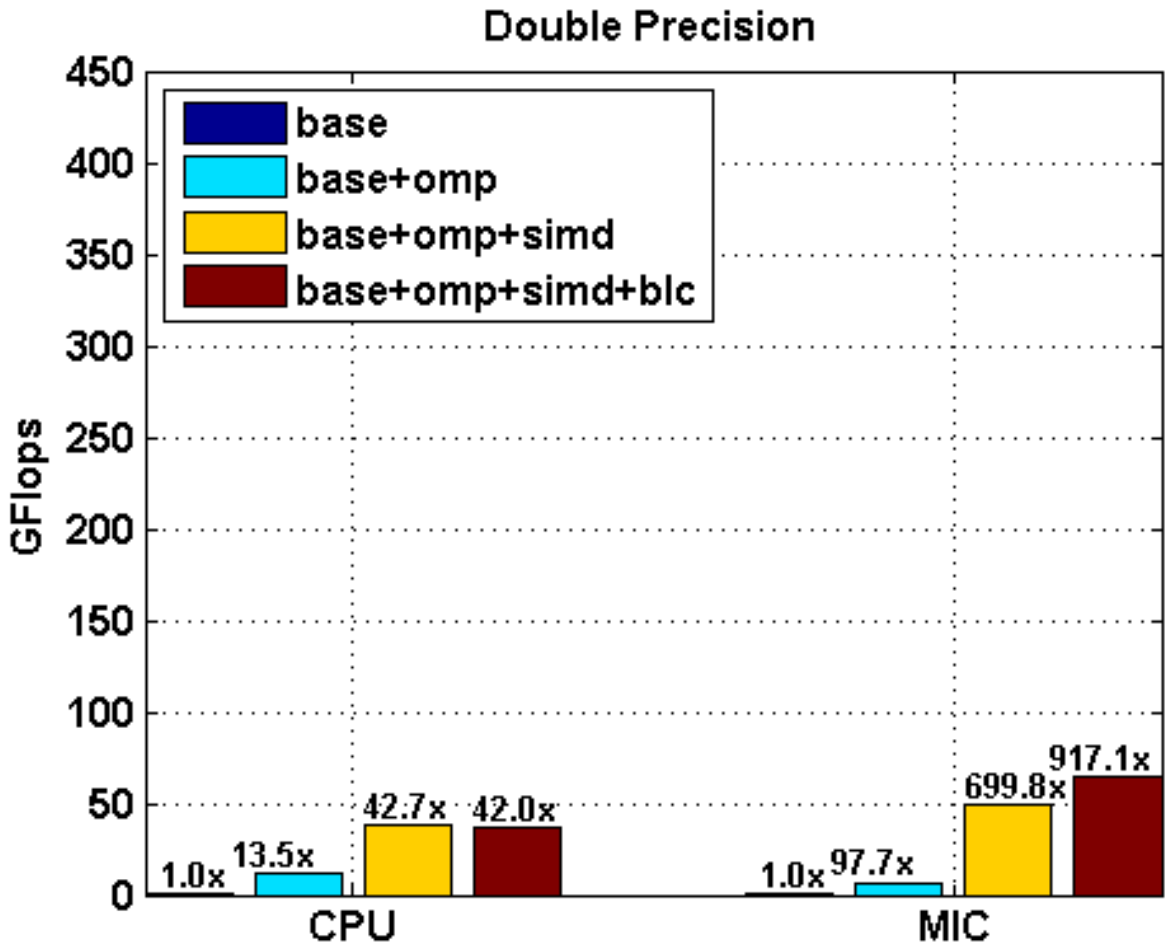


Note: C2M — data movement from CPU to MIC
M2C — data movement from MIC to CPU
halo1/2/3/4 — the 4 steps of the "pipe-flow" communication scheme

Optimization Scheme

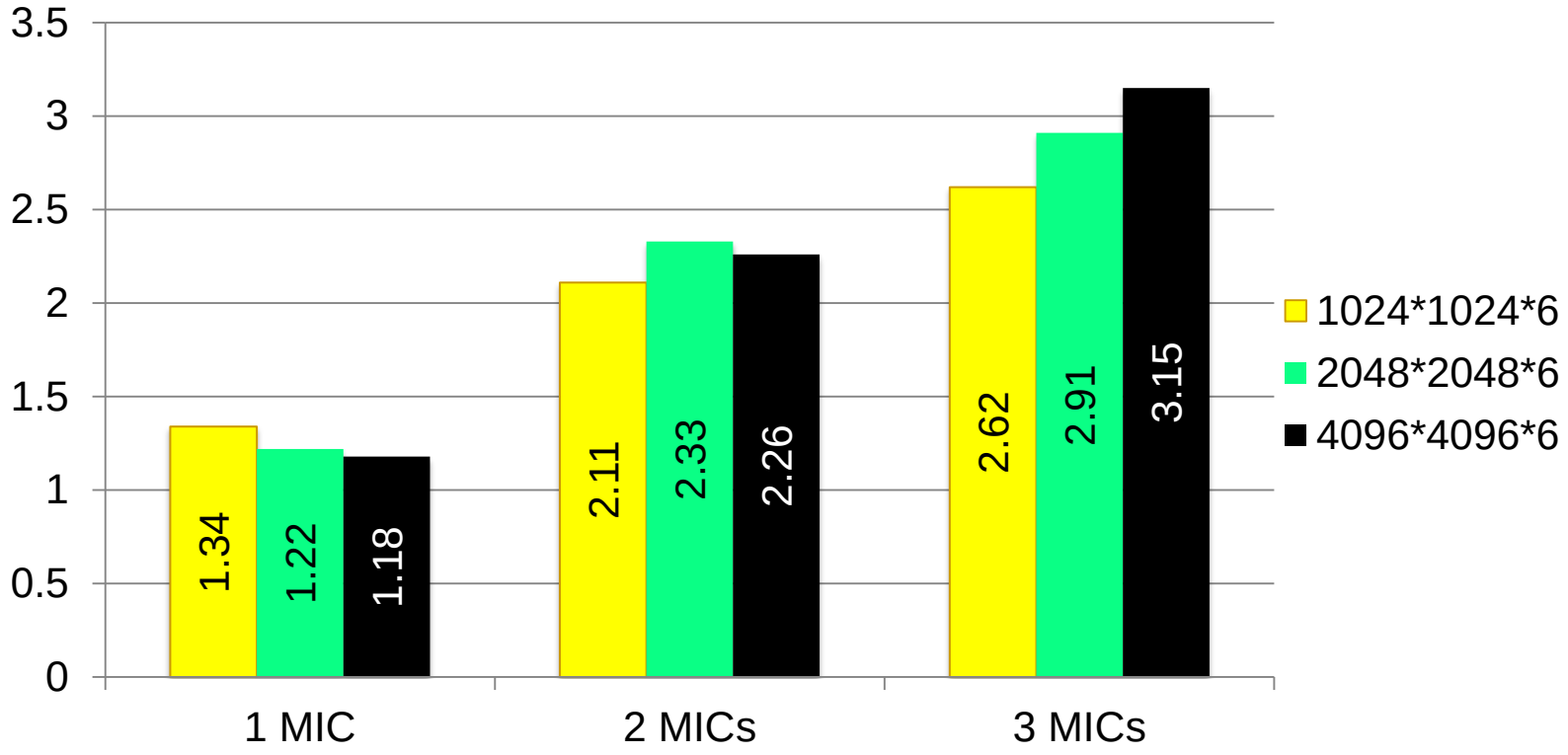


Scaling the Performance



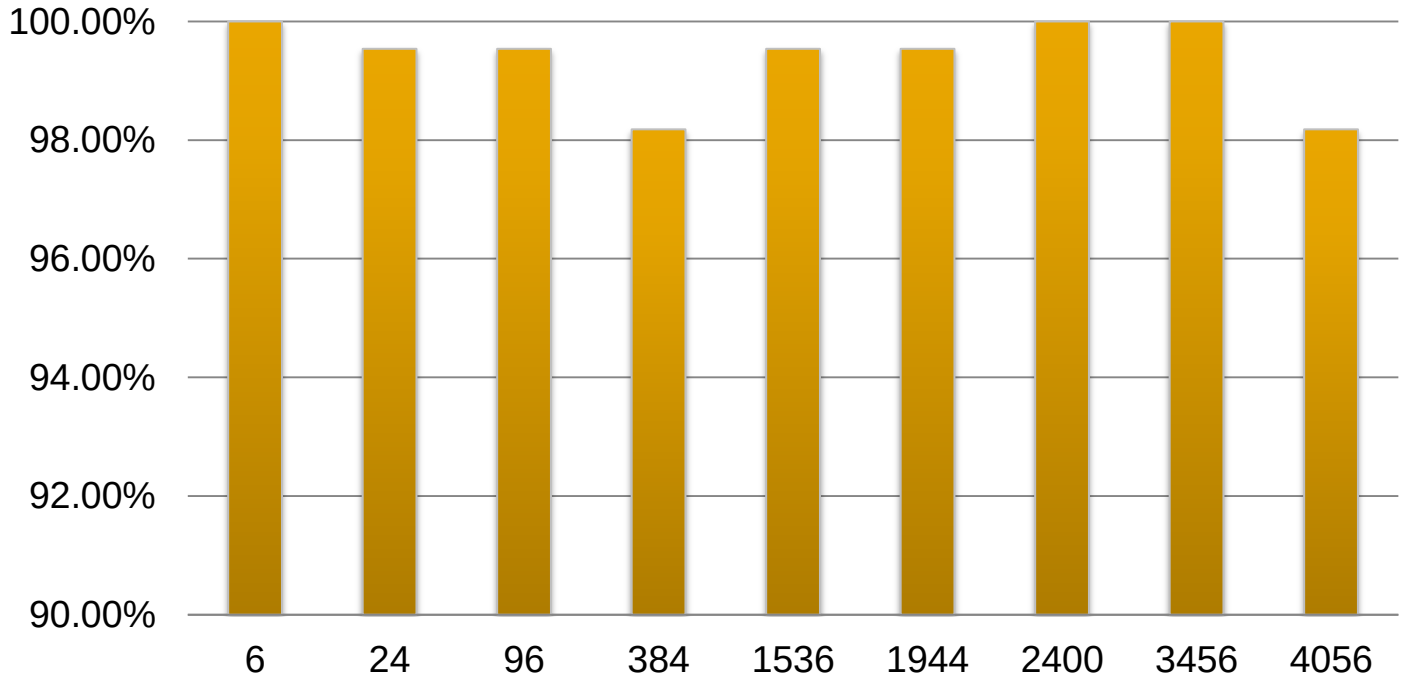
SWE Performance on Tianhe-2

- MIC against 24 CPU cores



SWE Performance on Tianhe-2

- Weak Scaling
  - Using up to 8,652 nodes (207,648 CPU cores + 1,583,316 MIC cores)



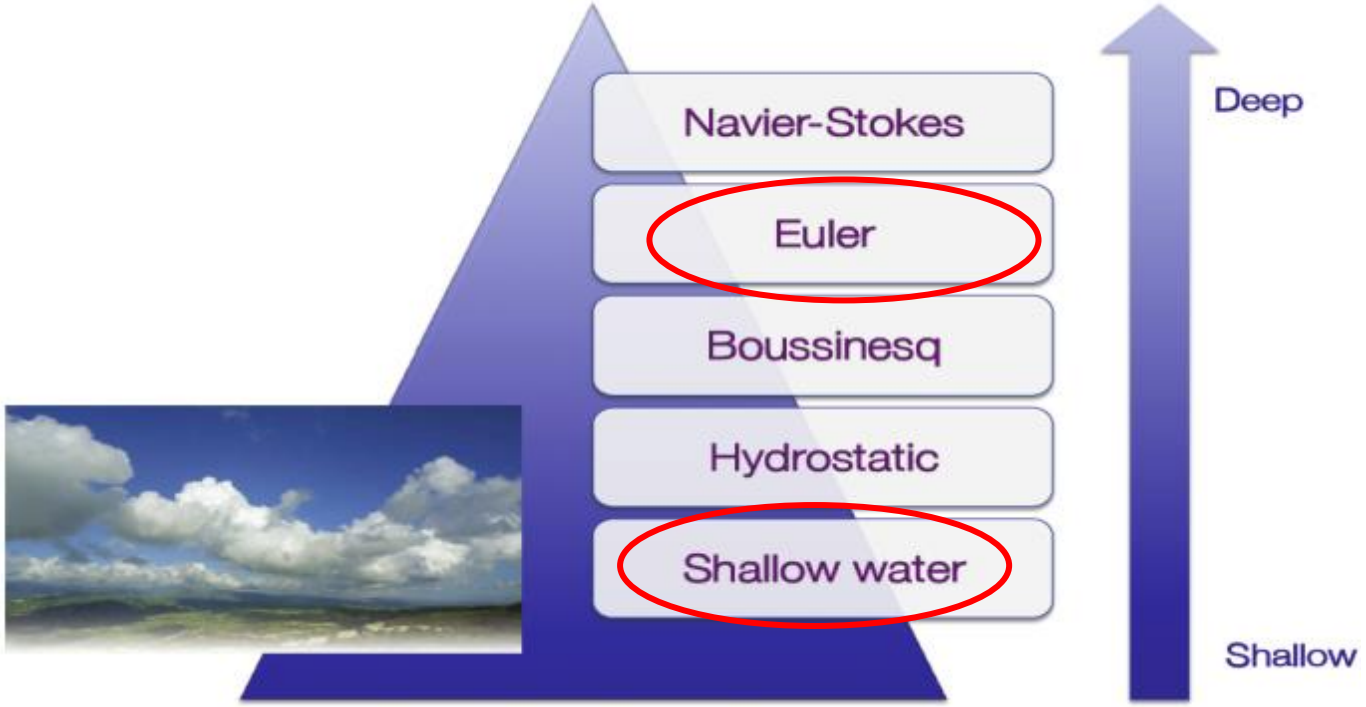
# of unknowns for the largest run: 200 billion

Outline

- Tianhe-1A: GPU
- Maxeler DFE: FPGA
- Tianhe-2: MIC
- Future Plan & Discussion

Highly-Scalable Framework for Global Atmospheric Simulation

- evolve from “2D Shallow Water Wave Equations” to “3D Euler Equations”



Highly-Scalable Framework for Global Atmospheric Simulation

- Model development:
  - from 2D SWE to 3D Euler
  - coupling the 3D Euler dynamics with physics processes to test for local and global scenarios
- HPC:
  - an FPGA-based cluster for climate modeling?
  - larger-scale runs on 100P supercomputer (dynamic + physics)

Thank You!

haohuan@tsinghua.edu.cn

A Practical Data Flow Design with 4866 nodes